

VYTAUTAS DUMBLIAUSKAS, Ph.D. student<sup>1</sup>

(Corresponding author)

E-mail: vytautas.dumbliauskas@vgtu.lt

VYTAUTAS GRIGONIS, Ph.D.<sup>1</sup>

E-mail: vytautas.grigonis@vgtu.lt

ANDRIUS BARAUSKAS, Ph.D. student<sup>1</sup>

E-mail: andrius.barauskas@vgtu.lt

<sup>1</sup> Vilnius Gediminas Technical University

Department of Roads

Saulėtekio avenue. 11, LT-10223 Vilnius, Lithuania

Traffic Management

Preliminary Communication

Submitted: 25 Dec. 2016

Accepted: 10 Oct. 2017

# APPLICATION OF GOOGLE-BASED DATA FOR TRAVEL TIME ANALYSIS: KAUNAS CITY CASE STUDY

## ABSTRACT

Recently, new traffic data sources have emerged raising new challenges and opportunities when applying novel methodologies. The purpose of this research is to analyse car travel time data collected from smartphones by Google Company. Geographic information system (GIS) tools and Python programming language were employed in this study to establish the initial framework as well as to automatically extract, analyse, and visualize data. The analysis resulted in the calculation of travel time fluctuation during the day, calculation of travel time variability and estimation of origin-destination (OD) skim matrices. Furthermore, we accomplished the accessibility analysis and provided recommendations for further research.

## KEY WORDS

Kaunas City; Google Traffic Data; Python; Skim Matrix;

## 1. INTRODUCTION

Most traffic engineering or transport planning tasks begin with a collection of traffic data. The major disadvantage of traditional traffic monitoring technologies (sensor-based, video-based, radio-frequency-based) is high deployment and maintenance cost. Additionally, these are also limited in coverage [1].

As an alternative approach, the collection of location-based positioning data was firstly accomplished by using cell phone signals between cell towers. Mobile-phone carriers automatically collect Call Detail Records (CDR) which contain time-stamped coordinates of anonymized customers. This information can provide detailed spatial-temporal information regarding user's mobility pattern [2], which can be used for origin-destination (OD) matrix development [2-7], estimation of travel times [8, 9], composition of traffic analysis zones [10]. However, the emerging Global Positioning System (GPS) technologies, embedded in smartphones, provide higher location data resolution and accuracy, while CDR technology usually only irregularly exchanges information with cell-towers,

GPS data are typically collected much more frequently (0.5–60 times per minute). According to researchers [11], GPS technology can provide 10-15 metres' accuracy (up to 5 metres in open areas), whereas CDR only 100 metres.

Apart from the high precision, smartphones with assisted GPS demonstrate high penetration rates and since almost everyone is using smartphone, it has become an effective sensor of our daily whereabouts [12].

The precise assisted GPS data might be employed to gain new innovative insights. For example, using GPS trace data from smartphones, the researchers [13, 14] have demonstrated how to reconstruct travel paths, which precisely match the road network geometry.

There are also studies on detecting transport modes. The travel mode classification algorithms, employing accelerometer and GPS data from mobile phones, has been developed and tested in real-life conditions [15-18]. Authors [15] claim that by using the right classification algorithm, one can determine the travel mode, such as walking, biking or using motorized transport as accurately as with 93.6%. Data gathering from smartphones is low-cost and can be done frequently.

Other authors [19] have also proposed a modelling framework that identifies both the physical paths and transport modes simultaneously, in order to take advantage of the correlation between them.

It can be seen, that there exists an enormous methodological potential, which can be applied to location datasets collected from mobile phones assisted by GPS. However, the availability of real data is limited and poses significant concerns, as privacy issues have not been properly addressed yet.

The purpose of this research is to promote freely available Google Maps travel time data collected from a large number of smartphones with built-in GPS sensors and showcase analysis methodology. In this

paper, we have demonstrated how to access data and to apply analysis techniques which are relevant in transport planning and modelling fields.

The researchers [20] have already attempted to make use of Google Maps data by estimating OD travel time matrices. Our research expands this study by extracting travel time for much more fine-grained zoning system and at several time moments during the day. This allowed us to successfully accomplish journey time variability analysis at the city scale. Furthermore, we have performed zone accessibility analysis, which may be relevant to local investors or city authority.

The whole paper is organized as follows: the first section establishes the object of analysis and spatial framework, the second section determines data sources and discusses privacy issues, the third and fourth sections present data extraction, processing, and visualisation techniques, the fifth section discusses implications of results, and the last section provides conclusions.

## 2. OBJECT OF ANALYSIS AND SPATIAL FRAMEWORK

The object of this study is Kaunas city and its surroundings (*Figure 1*) located in the central part of Lithuania. Based on the data of the Lithuanian Department of Statistics [21] in 2015 the city had 299.5 thousand residents and covered an area of 157 km<sup>2</sup>.

In order to establish initial data for further investigation and analysis, we divided the area in question into transportation analysis zones. The zones are

usually bounded by main streets, administrative or geographical boundaries, and include territories that possess functional integrity.

The area of study was actually bigger than the administrative boundaries of Kaunas city; therefore, all zones were differentiated based on their location as follows:

- Urban zones, located within administrative boundaries of the city;
- Suburban zones, located outside administrative boundaries of the city and within the area of interest;
- External zones, located on main roads outside the area of study. They represent “the rest of the World”.

Additionally, we assigned a centroid with a number to each zone. The centroid represents the mean centre of population living in that zone.

This zoning system is identical to the one which is usual in transport system macro modelling. Our approach followed the same rules because the zoning system and results of this study (OD skim matrix) are to be used for calibrating Kaunas city macro model.

The area of study covers 309 km<sup>2</sup>, 152 km<sup>2</sup> of which are suburban areas and 157 km<sup>2</sup> – urban areas. *Table 1* provides the summarized details of all zones.

As indicated in *Table 1*, there are 206 zones. In terms of the area in question, the suburban zones are around six times as big as the urban zones. Nonetheless, the number of residents in all the zones is approximately the same.



Figure 1 – A scheme of study area

Table 1 – Zone attributes

Zone type	Label in Figure 1	Number of zones	Average area [km <sup>2</sup> ]	Average number of residents
External zones	188 - 205	17	*	*
Suburban zones	165 - 188	24	6,35	1,970
Urban zones	0 - 164	165	0,95	2,065
Total	-	206	-	-

\* these data are unavailable as external zones are a hypothetical construct and do not have area or number of residents

### 3. DATA SOURCE: NATURE AND PRIVACY ISSUES

The data source used in this study was the Google Maps database, which provides access not only to the network geometry but also to the transport travel times. Google Maps bases its traffic travel time data on two types of information: historical data regarding the average travel time on a particular road section at a specific time on specific days and real-time data indicating real-time speed of vehicles.

How does the Google Company collect these data? Initial versions of Google Maps relied only on traffic sensor data, most of which were installed by government transportation agencies or private companies that specialized in traffic data compilation. Starting from 2009, Google turned to crowdsourcing to improve the accuracy of its traffic predictions. Crowdsourcing makes use of travel time data collected from smartphone users.

When Android smartphone users turn on their devices and enable their GPS, smartphones send back anonymous bits of location data to Google database. This data allows the company to identify movement trajectories, make further analysis and provide anonymised aggregated information (speeds and travel times) to wider audience. As more and more drivers use the application, aggregated information (travel times and speeds) have become more reliable [22].

Such crowdsourcing is very sensitive to location privacy issues. Location privacy, also known as geo-privacy, is a person's right to prevent the disclosure of their location data or to determine how and to what extent that information can be shared with others [23]. This means that raw data cannot be published, unless the user agrees to share it. This leads to dilemma of how to provide professionals with data access while preserving privacy. However, there exist studies [23], providing obfuscation methods, such as random perturbation and grid masking, which could be applied to GPS trajectory data in order to set balance between privacy concerns and spatial pattern disruption. Some authors [24] suggest an alternative solution: to let only approved users through secure portal environment work with detailed spatial data.

So, Google company allows only anonymised and aggregated data to be accessed via application programming interface, namely Google Maps Distance Matrix API. The procedure used to extract data is described in the following section.

### 4. DATA EXTRACTION

Google Maps Distance Matrix API is a service that returns travel distance and time for an origin-destination (OD) pair at a specified departure time. The information received is based on the recommended route between origin and destination, as calculated by the Google Maps Distance Matrix API [25].

In our analysis, there were number of OD pairs, where  $n^2 - n$  is the number of transport analysis zones (see section No. 2). Additionally, we extracted each OD pairs' travel time at thirty-minute intervals during the day (from 3:00 to 21:30), which resulted in slightly more than 1.6 million requests.

Every request was sent by a script, which constructed the Uniform Resource Locator (URL) by using the following parameters [25]:

- Longitude of origin (required);
- Latitude of origin (required);
- Longitude of destination (required);
- Latitude of destination (required);
- Mode of travel (optional);
- Date and time of departure (optional).

Latitudes and longitudes geocoded the centroids of OD pairs and were variable in each request. Mode of travel was kept constant ('driving') as further analysis is restricted to the motorized private transport travel times on the urban street network. We also restricted our travel time analysis to weekdays starting at 3:00 and ending at 21:30 (with 38 intermediate points every half an hour).

Results of Google Distance Matrix API were provided as JavaScript Object Notation (JSON) objects, an open-standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs. Every answer to the request (JSON object) contained the following root elements:

- Origin addresses;
- Destination addresses;
- Additional rows which indicated duration and distance in metres and seconds respectively.

Unfortunately, Google Distance Matrix API provides no additional metadata about the result returned to the user. It is unknown what sample size was used to calculate average travel times, so the reliability remains unclear and cannot be estimated. Having said that, it should be noted that in 2015 Lithuania's smartphone penetration rate was 42%, so it is likely that reliability issue might not be significant.

Scripting implementation and the results of all the requests can be provided for interested readers upon request.

## 5. DATA PROCESSING AND VISUALISATION

Data extracted from Google Distance Matrix API comprise a large dataset composed of more than 1.6 million rows. Therefore, comprehensive tools were required to accomplish the analysis and visualisation efficiently.

Python library called NumPy was considered to be a suitable tool to process the data. NumPy is a general-purpose array-processing package designed to efficiently manipulate large multi-dimensional arrays of arbitrary records without sacrificing too much speed for small multi-dimensional arrays [26].

Furthermore, Python library Seaborn was chosen to communicate the results of analysis. Seaborn is a Python visualization library which provides a high-level interface for drawing attractive statistical graphics [27].

Visualisation of results with spatial context requires application of Geographic Information System. This was done using an open source geographic information system QGIS. This system is a user-friendly GIS that provides common functions and features. The software has developed in a way that it is currently used to satisfy the daily GIS data-viewing needs. QGIS supports a number of raster and vector data formats and a new format support is easily added using the plug-in architecture [28].

### 5.1 Data processing

To complete visualisation, initial data needed to be saved in a universal format. This was done by using three-dimensional NumPy array with  $n$  rows,  $m$  columns and  $s$  layers (Figure 2). A cell, where  $i$ -th row and  $j$ -th column intersect, contains travel time from  $i$ -th zone to  $j$ -th zone. The third  $k$ -th axis signifies departure moments, which vary from 3:00 to 21:30 in intervals of 30 minutes (38 intervals in total). All cells of the array were filled with values of weekdays travel time, which varies among pairs of zones and during different periods of day due to the fact that changing travel demand influences travel time.

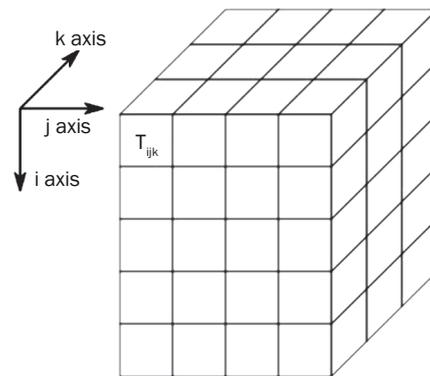


Figure 2 – An array representing travel times

Once two three-dimensional array is constructed, we can proceed with data analysis. The following sections provide our results, namely:

- a graph of weekdays relative travel times,
- a matrix of mean weekday travel times,
- a matrix of travel time coefficient of variation,
- a graph of coefficient of variation and travel times,
- a matrix of relative evening peak delays,
- a scheme of accessibility.

### 5.2 Relative travel times

Relative travel times of weekdays were calculated as follows: for every two-dimensional array along  $k$  axis, all travel time values located at  $i$  and  $j$  pairs were summed up according to the following formula:

$$T_k = \sum_{i=1}^{i=n} \sum_{j=1}^{j=m} T_{ijk} \quad (1)$$

Then the resulting vector values were divided by the maximum element of the same vector. The results of the analysis are shown in Figure 3.

Based on the graph, all daytime travel times can be conceptually divided into four periods: night time, morning rush, off-peak, and evening rush. The morning rush period reached its peak at 7:30 am, while the evening rush – at 17:00. Furthermore, the morning peak congestion is less intensive than the evening by 4%, even though visually they are shaped similarly. Off-peak travel times (between 9:30 and 15:30) are about 10% lower than the evening peak. Additionally, night time period travel times are only 15% lower than the evening peak.

### 5.3 Mean travel time variability during typical day

Further, we estimated travel time variability, which is expressed with a simple statistical measure, such as coefficient of variation. The coefficient of variation is defined as the ratio of the standard deviation to the mean:

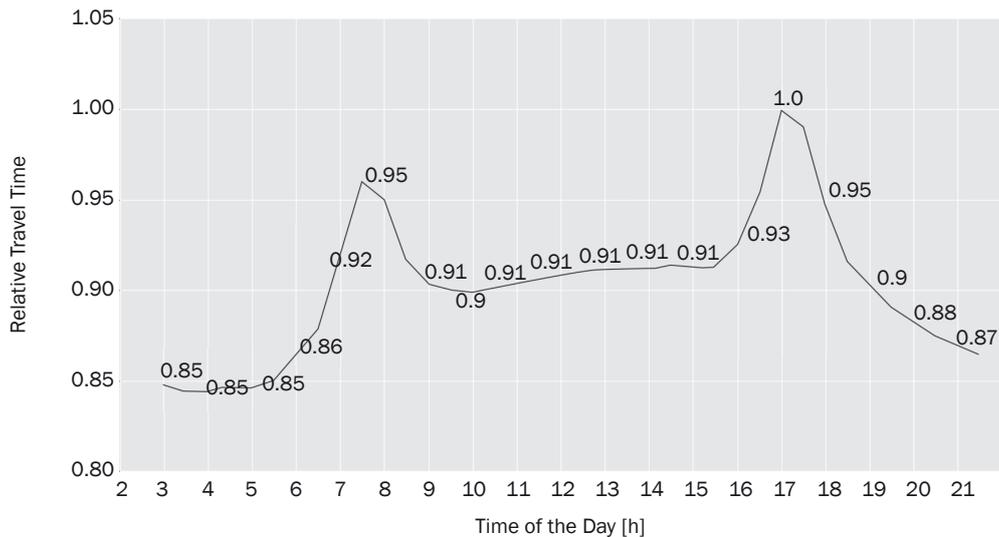


Figure 3 – Weekdays relative travel times

$$CVAR_{ij} = \frac{STD_{ij}}{\bar{T}_{ij}} \quad (2)$$

where:

$\bar{T}_{ij}$  - is the mean travel time between origin  $i$  and destination  $j$ ;

$STD_{ij}$  - is standard deviation of travel time between origin  $i$  and destination  $j$ .

Standard deviation between all origin and destination pairs is calculated as follows:

$$STD_{ij} = \sqrt{\frac{1}{s} \sum_{k=1}^{k=s} (T_{ijk} - \bar{T}_{ij})^2} \quad (3)$$

$T_{ijk}$  - is the extracted car travel time between origin  $i$  and destination  $j$  at moment  $k$ ;

$s$  - number of moments at which car travel times were extracted.

Finally, we plot the coefficient of variation against the travel times and the results are shown in Figure 4.

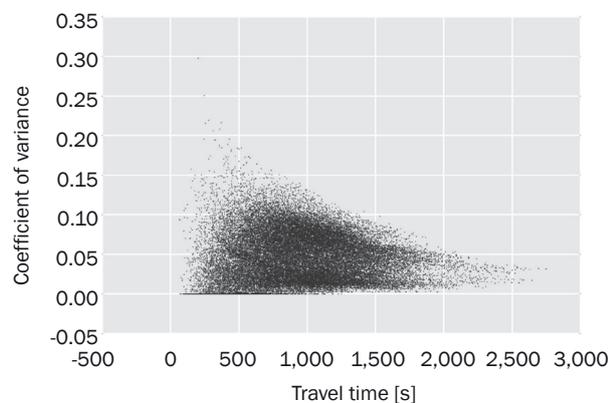


Figure 4 – Relative travel times

As can be seen from the scatterplot, the range across which coefficient of variation fluctuates highly depends on the travel time itself. The shorter the road

section, the wider is the range in which coefficient of variation may fall. This means that variability of the coefficient of variation is higher for shorter travel time sections and lower for longer ones. Analogously, we can identify that coefficient of variation continuously decrease as travel times become larger and larger.

Transport studies occasionally require travel time surveys along predefined routes. Multiple journey time observations are necessary for a route, in order to show a required accuracy, which is usually  $\pm 10\%$ , with 95% confidence level.

This accuracy level and underlying travel time variability along the route will determine the required sample size of observed journey times. However, journey time variability (variance or standard deviation) is often not known and should be assumed, which sometimes result in a necessity to make more than one site visit, as observed variability appears to be higher than assumed. A provided variability analysis may assist engineers in their assumptions about travel time variability and help them to avoid repeated site visits.

### 5.4 Matrix of mean travel times

The matrix of travel times (in technical terms - skim matrix) in practise is usually calculated by macro-modelling software. Estimation of its values is based on paths, which are calculated during the assignment procedure. Our approach makes use of data collected from smartphones with GPS and therefore it reflects the reality with utmost confidence. The values of the matrix were calculated as follows:

$$\bar{T}_{ij} = \frac{1}{s} \sum_{k=1}^{k=s} T_{ijk} \quad (4)$$

Figure 5 shows travel times averaged over all 38 periods during the day.

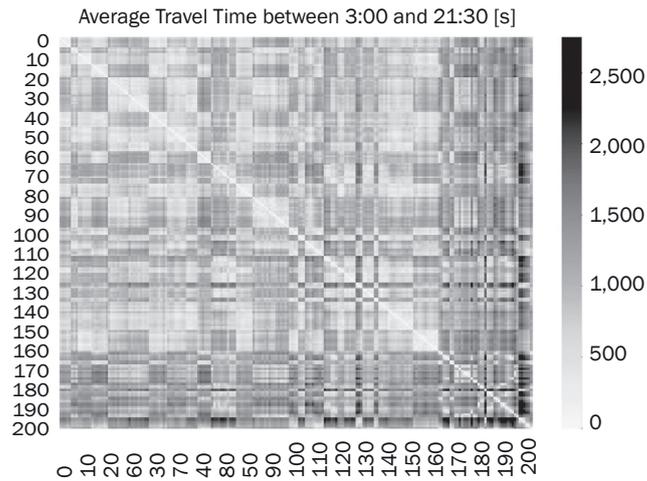


Figure 5 – Average weekdays travel time matrix

Each value is indicated by a colour scheme provided on the right-hand side. Elements of the matrix diagonal are equal to zero, while other values vary between 0 and 3,000 seconds. Visual inspection of the matrix reveals that most of the zones with numbers over 164 are displayed in dark colours, meaning longer travel times. This is indeed the case, as the indicated zones are suburban zones located outside the city administrative area. Urban zones feature lower travel times as light colour is dominating in these cells.

This matrix identifies mean travel time values, which were obtained by averaging travel times during 38 periods, however it is very easy to produce such a matrix for any particular period. For example, morning or evening peak matrices may be created and used to validate the city macromodel.

### 5.5 Accessibility of transport analysis zones

Similarly to other authors [20] the collected data also give us an opportunity to evaluate the accessibility of each zone. This was done by calculating the friction index based on the evening peak travel time matrix. Firstly, we established the total departing time vector by adding values along the rows of evening peak matrix. Secondly, we calculated the arriving time vector by adding values along the columns. The elements of the first vector represented the total travel time required to reach all destinations from a corresponding zone. The elements of the second vector represented the total travel time required to arrive to a corresponding zone from all other places of origin. Then both vectors were summed up, element by element. Finally, the acquired vector was divided by its

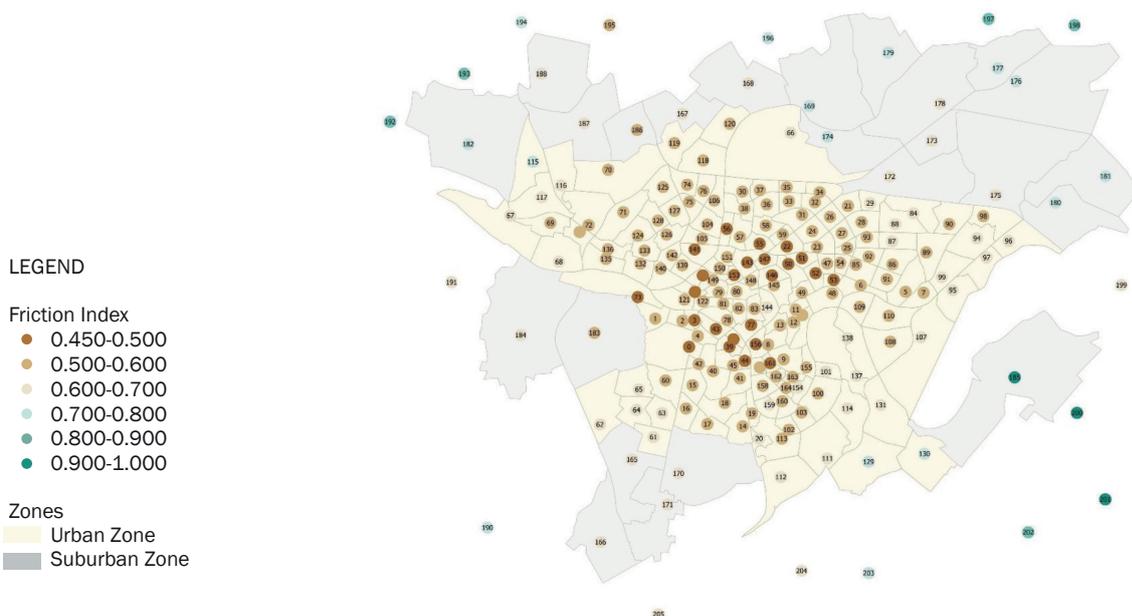


Figure 6 – Accessibility of centroids

maximum value. Due to this method the friction index was valued between 0 and 1. The higher the friction index, the lower the accessibility. The results of the analysis are provided in *Figure 6*.

Centroids with low friction index and the best accessibility are indicated in dark brown colour. There are few clusters of high accessibility centroids and these are located along the most important arterial streets in the central part of the city, namely: Taikos Avenue, Savanoriu Avenue and Karaliaus Mindaugo Avenue. In contrast, the worst accessibility is attributed to the remote external zones located to the east-south of the city. Our local knowledge verifies this result, as this area is poorly connected to the core urban transport network.

This scheme can be used for various purposes. First, it is helpful for potential investors in their location decision-making process, as they seek to acquire land with the best accessibility and the lowest purchase price. Furthermore, government agencies and planners can examine the causes of poor accessibility and provide measures for improvement. A detailed examination of the network is not a subject of this study.

## 6. DISCUSSION

Over the past few years the researchers have observed new traffic data sources. The most prominent is the location and movement trajectory GPS data implemented by main mobile operating system developers, such as Google Android and Apple iOS. Advantages of mobile sensing, in comparison with fixed-location sensing (e.g. using loop detectors and cameras), include a potentially complete spatial and temporal coverage of traffic network and high positioning accuracy [29]. Moreover, the mobile sensing system is advantageous as it does not need to be installed separately and it barely incurs any maintenance cost.

Mobile sensing system itself has seen significant development during last decades and with an introduction of precise assisted GPS it generates a huge amount of data that could be employed in transport planning and engineering field. However, due to the privacy issues the potential of these data cannot be used as the owners of these data (for example Google or Apple companies) are bound to keep it secretly and are not willing to reveal raw movement GPS trajectories.

Google company, nevertheless, provides aggregated travel time data, which can be accessed via Google Distance Matrix API. Unfortunately, there is no information about the sample sizes employed to calculate the mean value, thus making the data reliability unknown.

During the course of this study, we calculated car travel time fluctuation during the weekday, estimated travel time variability and origin-destination (OD) skim matrices. Furthermore, the accessibility analysis was performed.

The travel time fluctuation appeared to be very characteristic and verified the current knowledge about demand patterns in Kaunas city. Not surprisingly, it revealed two peak periods, which reached their highest points at 7:30 and 17:00. The evening peak period was approximately 4% heavier in congestion compared to the morning peak period. No midday peak period was identified. Similar analysis could have been accomplished also for any particular OD pair or even short travel time section (described by two closely located OD points). The applicability might be extended to various cases.

Car travel time variability analysis revealed that the range of the coefficient of variation is wider for shorter travel time sections and narrower for longer ones. It was also clear that with longer travel time sections the coefficient of variation decreases continuously. The results of the analysis may be used like reference material as car travel time variability is often found as input into other tasks, for example, determination of the number of observations for car travel time surveys.

OD travel time matrices provide useful data, which can also be used in other applications, for example, to validate city-wide strategic model. In such a case, we would calculate peak time OD travel time matrix corresponding to the period being modelled.

The accessibility analysis allowed us to evaluate network connectivity and this result may prove to be very useful for investors, as the zones with good accessibility are very attractive locations for industrial or business developments.

The methodology used can also be applied in future research. For example, we can estimate OD travel time matrices in different seasons of the year to account for the fluctuating weather conditions throughout the year, or to estimate OD travel time matrices (or accessibility vector) before and after the implementation of major network improvements (bypasses) or restrictions (congestion charges).

## 7. CONCLUSIONS

Google company provides rich set of data on car travel times which can be accessed via Google Distance Matrix API. By using open source tools (Python programming language and its libraries, QGIS), we were able to establish the spatial framework and extract data for the whole Kaunas city and its surroundings.

Our analysis firstly resulted in the graph of relative travel times. Later, we looked into travel time variability. We also produced the mean travel time matrix which

might be used as a skim matrix to validate the macro model of Kaunas city. Finally, we concluded with the evaluation of accessibility of all 206 centroids.

All data were effectively processed and visualised using Python programming language and its libraries. Python is a general-purpose programming language, which is rich in scientific libraries, such as Numpy, Scipy, Matplotlib, or Seaborn. Moreover, all the tools used in this research were open source and did not require any investment in commercial software.

**VYTAUTAS DUMBLIAUSKAS**, doktorantas<sup>1</sup>

E-mail: vytautas.dumbliauskas@vgtu.lt

**VYTAUTAS GRIGONIS**, docentas<sup>1</sup>

E-mail: vytautas.grigonis@vgtu.lt

**ANDRIUS BARAUSKAS**, doktorantas<sup>1</sup>

E-mail: andrius.barauskas@vgtu.lt

<sup>1</sup> Vilniaus Gedimino technikos universitetas, Kelių katedra  
Saulėtekio al. 11, LT-10223 Vilnius, Lietuva

## GOOGLE KOMPANIJOS TEIKIAMŲ KELIONĖS LAIKO DUOMENŲ ANALIZĖ: KAUNO MIESTO ATVEJO STUDIJA

### SANTRAUKA

*Pastaruoju metu eismo inžinerijos ir planavimo srityje labai aktyviai formuojasi nauji duomenų šaltiniai, kurie sukuria poreikius ir inovatyviems jų apdorojimo būdams. Šio tyrimo tikslas yra sukurti Google kompanijos teikiamų kelionės laiko duomenų analizės metodiką. Automatizuotam duomenų išgavimui, analizei ir grafiniam atvaizdavimui, tyrimo metu buvo pritaikytos geografinės informacinės sistemos (GIS) ir Python programavimo kalba. Analizės metu buvo identifi-kuotas kelionės laikas tarp visų miesto transportinių rajonų porų, įvertinant ir jo svyravimus paros bėgyje. Galiausiai atliktą ir miesto transportinių rajonų pasiekiamumo analizę, kuri yra naudinga tiek viešojo sektoriaus atstovams vystant transporto sistema, tiek privačiam sektoriui planuojant investicijas. Straipsnio pabaigoje pateikiame rekomendacijas tolimesniam metodikos vystymui.*

### RAKTINIAI ŽODŽIAI

*Kauno miestas; Google duomenys; Python; kelionių laiko matrica;*

### REFERENCES

- [1] Wang F, Hu L, Zhou D, Sun R, Hu J, Zhao K. Estimating online vacancies in real-time road traffic monitoring with traffic sensor data stream. *Ad Hoc Networks*. 2015;35: 3-13.
- [2] Alexander L, Jiang S, Murga, M, González M C. Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*. 2015;58: 240-250.
- [3] Iqbal MS, Choudhury CF, Wang P, González MC. Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*. 2014;40: 63-74.
- [4] Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, Gonzalez MC. The path most travelled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*. 2015;58: 162-177.
- [5] Calabrese F, Di Lorenzo G, Liu L, Ratti C. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Computing*. 2011;10(4): 36-44.
- [6] Widhalm P, Yang Y, Ulm M, Athavale S, Gonzalez MC. Discovering urban activity patterns in cell phone data. *Transportation*. 2015;42(4): 597-623.
- [7] Çolak S, Alexander LP, Alvim BG, Mehndiratta SR, González MC. Analysing Cell Phone Location Data for Urban Travel. *Transportation Research Record: Journal of the Transportation Research Board*. 2015;2526: 126-135.
- [8] Bar-Gera H. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies*. 2007;15(6): 380-391.
- [9] Lovell DJ. Accuracy of Speed Measurements from Cellular Phone Vehicle Location Systems. *ITS Journal - Intelligent Transportation Systems Journal*. 2001;6(4): 303-325.
- [10] Dong H, Wu M, Ding X, Chu L, Jia L, Qin Y, Zhou X. Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C: Emerging Technologies*. 2015;58: 278-291.
- [11] Patire AD, Wright M, Prodhomme B, Bayen AM. How much GPS data do we need? *Transportation Research Part C: Emerging Technologies*. 2014;58: 325-342.
- [12] Lane N, Miluzzo E, Lu H. A Survey of Mobile Phone Sensing. *IEEE Communications Magazine*. 2010;48(9): 140-150.
- [13] Bierlaire M, Chen J, Newman J. A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*. 2013;26: 78-98.
- [14] Goh C, Dauwels J, Mitrovic N. Online Map-Matching based on Hidden Markov Model for Real-Time Traffic Sensing Applications. *The 15<sup>th</sup> International IEEE Conference on Intelligent Transportation Systems*. 16-19 Sep. 2012, Anchorage, AK, USA. IEEE; 2012. p. 776-781.
- [15] Reddy S, Mun M, Burke J, Estrin D, Hansen M, Srivastava M. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks*. 2010;6(2): 1-27.
- [16] Widhalm P, Nitsche P, Brändle N. Transport Mode Detection with Realistic Smartphone Sensor Data. *21st International Conference on Pattern Recognition (ICPR 2012)*, 11-15 Nov. 2012, Tsukuba, Japan. IEEE; 2012. p. 573-576.
- [17] Bohte W, Maat K. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*. 2009;17(3): 285-297.
- [18] Nitsche P, Widhalm P, Breuss S, Maurer P. A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys. *Procedia - Social and Behavioral Sciences*. 2012;48: 1033-1046.
- [19] Chen J, Bierlaire M. Probabilistic Multimodal Map Matching with Rich Smartphone Data. *Journal of Intel-*

- ligent Transportation Systems. 2015;19(2): 134-148.
- [20] Wang F, Xu Y. Estimating O-D travel time matrix by Google Maps API: implementation, advantages, and implications. *Annals of GIS*. 2011;17(4): 199-209.
- [21] The Department of Statistics. Official Statistics Portal. 2015. Available from: [osp.stat.gov.lt](http://osp.stat.gov.lt)
- [22] Barth D. The Bright Side of Sitting in Traffic: Crowdsourcing Road Congestion Data. 2009. Available from: <https://googleblog.blogspot.co.uk/2009/08/bright-side-of-sitting-in-traffic.html>
- [23] Seidl DE, Jankowski P, Tsou MH. Privacy and spatial pattern preservation in masked GPS trajectory data. *International Journal of Geographical Information Science*. 2015;30(4): 1-16.
- [24] Gonder J, Burton E, Murakami E. Archiving Data from New Survey Technologies: Enabling Research with High-precision Data While Preserving Participant Privacy. *Transportation Research Procedia*. 2015;11: 85-97.
- [25] Google. Google Maps Distance Matrix API. Developer's Guide. Available from: <https://developers.google.com/maps/documentation/distance-matrix/intro#Introduction>
- [26] The Python Software Foundation. NumPy: Array Processing for Numbers, Strings, Records, and Objects. Available from: <https://pypi.python.org/pypi/numpy>
- [27] Waskom M. Seaborn: Statistical Data Visualization. 2017. Available from: <https://stanford.edu/~mwaskom/software/seaborn/#>
- [28] QGIS. QGIS User Guide. Available from: <http://docs.qgis.org/2.8/en/docs/index.html>
- [29] Herrera JC, Work DB, Herring R, Ban X, Jacobson Q, Bayen AM. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*. 2010;18(4): 568-583.